Optimal Biorthogonal Filter Banks for Multiple Description Coding

Pier Luigi Dragotti[1] and Martin Vetterli[1,2]
[1] Dept. of Electrical Engineering, Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
[2] EECS Dept., UC Berkeley CA 947020
e-mail: {dragotti,vetterli}@lcavsun1.epfl.ch
http://lcavwww.epfl.ch

***Abstract* — Multiple description coding (MDC) is a source coding technique for information transmission over unreliable networks. In MDC, the coder generates several different descriptions of the same signal and the decoder can produce a useful reconstruction of the source with any received subset of these descriptions. This paper addresses the design of a two-channel biorthogonal filter bank for optimal MDC of wide sense stationary Gaussian input processes. The problem is solved in the frequency domain, the redundancy rate-distortion curve is constructed and, for each given redundancy, filter responses are designed to meet the corresponding point in the redundancy rate-distortion curve. An application of the proposed algorithm to a first order Gauss-Markov process is presented and for this case also approximated FIR solutions are proposed.**

I. INTRODUCTION

Recently the problem of transmitting data over unreliable packet switched networks has received considerable attention. Packet losses can be due to transmission errors or congestion. If the network is able to provide preferential treatment to some packets, then the use of multiresolution or layered source coding system is the obvious solution. But if the network cannot discriminate packets and retransmission is not allowed (due, for instance, to a delay constraint), the source coding strategy should be different. The transmitter should generate different descriptions of the source and put each of them in a different packet, so that, disregarding which subsets of these packets reach the receiver, it would be always possible to get an acceptable reconstruction of the source.

This problem is the generalization of the "multiple description problem" depicted in Fig.1. Here the transmitter generates only two different coded bitstreams (descriptions) of the source and sends them over two different erasure channel. If both descriptions are received then the decoder can reconstruct the source with a distortion D_0 (central distortion). If one of the bitstreams gets lost in the transmission, a degraded but still useful reconstruction quality can be achieved (D_1, D_2). In this second case, since the two descriptions have been previously correlated, the decoder can use the received bitstream to estimate the lost one based on the introduced statistical correlation. On the other hand this correlation reduces the coding efficiency. The excess rate ρ due to this suboptimality is called redundancy and represents the price we pay to have a robust transmission.

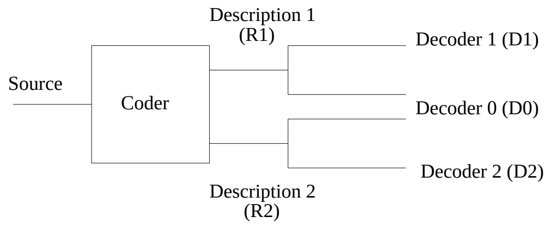

Fig. 1: The multiple description problem

Early papers on MDC are information theoretic and try to find the set of achievable values for the quintuple $(R_1, R_2, D_0, D_1, D_2)$ [1, 2]. More recent papers consider the problem of designing practical multiple description systems [3, 4, 5]. In [4, 5] a blockwise transform is applied to the input vector to obtain the multiple description property. The input vector is usually a jointly Gaussian vector, the basic idea being to decorrelate the vector components and then to introduce again correlation between coefficients but in a known and controlled manner so that erased coefficients can be statistically estimated from the received ones. In this paper we investigate the more general case of arbitrary and stationary Gaussian input process and two-channel biorthogonal filter banks. A first attempt to solve this problem was presented in [6], but the filter bank was constrained to be orthogonal; this kind of limitation turns out to be too restrictive since the multiple description property, at least in the blockwise transform context, is usually met using biorthogonal transforms. Our approach is similar to the one in the blockwise transform context. We construct a first set of filter banks to decorrelate the two input sequences and then we use a second set of filter banks to efficiently recorrelate them. The frequency response of this second set of filters depends on the amount of redundancy. The more redundancy we have the more correlated the two output sequences are. The proposed optimization algorithm is then applied to a first order Gauss-Markov process, but since the resulting filters are infinite length filters, we approximate them with FIR filters. We show that even FIR filters of moderate length can pretty well approximate the ideal behavior.

II. PROBLEM FORMULATION AND NOTATION

Consider the classical two-channel filter banks scheme shown in Fig.2. Here the input $x[n]$ is assumed to be a stationary Gaussian random process with known statistics and is passed through an analysis filter banks. The two output sequences are then separately quantized and sent over two different erasure channel. We suppose

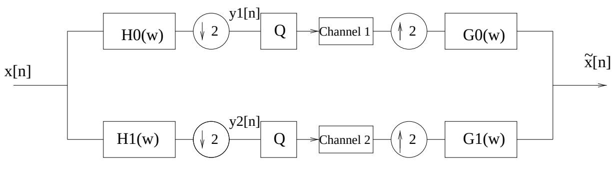

Fig. 2: Two channel filter banks

that the channels are independent and have the same erasure probability. For convenience we will formulate our problem in the polyphase domain [7]. In this case the analysis stage can be equivalently represented by the block scheme shown in Fig 3. The input-output relation can be expressed in matrix notation introducing the analysis polyphase matrix $H(\omega)$:

$$\begin{pmatrix} Y_1(\omega) \\ Y_2(\omega) \end{pmatrix} = \begin{pmatrix} H_{11}(\omega) & H_{12}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) \end{pmatrix} \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \end{pmatrix}. \tag{1}$$

$R_x(\omega)$ is the 2×2 polyphase power spectral density (p.s.d) matrix of the input process, so $R_{xij}(\omega)$ is the auto or cross p.s.d. between the ith and jth polyphase components. Likewise $R_y(\omega)$ is the p.s.d. matrix of the outputs, but here $R_{yij}(\omega)$ is the p.s.d. between the ith and jth channel signals. The system response has the following form:

$$R_y = H(\omega)R_x(\omega)H^+(\omega), \tag{2}$$

where $H^+(\omega)$ denotes the Hermitian transpose of $H(\omega)$. The synthesis part of the system can be analyzed in a similar fashion. Recall that given the analysis matrix, the synthesis polyphase matrix $G(\omega)$ is univocally defined (up to a phase factor). In fact $G(\omega)$ must be such that the condition $G(\omega)H(\omega) = I$ is verified [7].

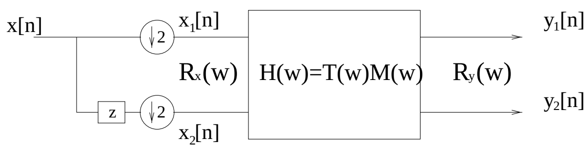

Fig. 3: The equivalent polyphase representation of the analysis stage

As anticipated in the previous section, the polyphase analysis matrix has to fulfill two tasks: first decorrelate the two input sequences, second efficiently recorrelate them according to the amount of available redundancy. For this reason we decompose the matrix $H(\omega)$ into the product of two matrices: $M(\omega)$ and $T(\omega)$.

$$H(\omega) = T(\omega)M(\omega). \tag{3}$$

$M(\omega)$ is the decorrelating matrix and its frequency response depends only on the statistics of the input signal; $T(\omega)$ is the "recorrelation" matrix and its frequency response depends on the redundancy; this is the matrix that, given the redundancy, has to be optimized. We call $R_{\hat{x}}$ the p.s.d matrix of the input process after decorrelation, clearly this matrix is diagonal. (See also Fig 4)

$$R_{\hat{x}} = \begin{bmatrix} \sigma_1^2(\omega) & 0 \\ 0 & \sigma_2^2(\omega). \end{bmatrix} \tag{4}$$

Now, considering that both channels are coded indepen-

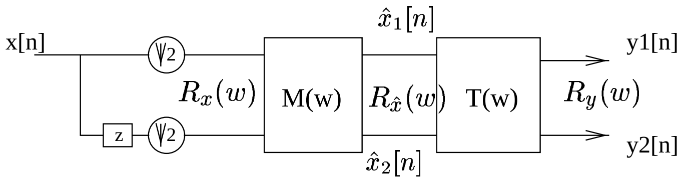

Fig. 4: The polyphase matrix decomposition into $M(\omega)$ and $T(\omega)$

dently, in terms of bit rate efficiency (without considering that the channels are unreliable) it would be better to code the two decorrelated sequences $(\hat{x}_1[n], \hat{x}_2[n])$ instead of the output sequences $(y_1[n], y_2[n])$. If our target central distortion is D_0 and if quantization is fine, the minimum bitrates for coding the two sequences $\hat{x}_1[n], \hat{x}_2[n]$ are given by the following formulas [8]:

$$\begin{aligned} \hat{R}_1(D_0) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} log \frac{\sigma_1^2(\omega)}{D_0} d\omega \\ \hat{R}_2(D_0) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} log \frac{\sigma_2^2(\omega)}{D_0} d\omega \end{aligned} \tag{5}$$

On the other hand the actual bitrates are the ones associated to the output sequences $y_1[n], y_2[n]$:

$$\begin{aligned} R_1(D_0) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} log \frac{R_{y11}(\omega)}{D_0} d\omega \\ R_2(D_0) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} log \frac{R_{y22}(\omega)}{D_0} d\omega \end{aligned} \tag{6}$$

We define as redundancy the difference rate between these two cases:

$$\begin{aligned} \rho &= (R_1(D_0) + R_2(D_0)) - (\hat{R}_1(D_0) + \hat{R}_2(D_0)) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} log \frac{R_{y11}(\omega) R_{y22}(\omega)}{\sigma_1^2(\omega)\sigma_2^2(\omega)}. \end{aligned} \tag{7}$$

Notice that ρ does not depend on D_0.

Now consider the case when one channel (i.e. channel 1) is cut off, then $y_1[n]$ must be estimated from the received sequence $y_2[n]$. The optimal estimation is obtained by Wiener filtering,

$$\hat{Y}_1(\omega) = \frac{R_{y12}(\omega)}{R_{y22}(\omega)} Y_2(\omega). \tag{8}$$

However, because we have used a nonorthogonal transform, we must return to the original space using $G(\omega)$ (which depends on $H(\omega)$) in order to compute the distortion. The final mean square error distortion D_2, omitting the quantization error, is:

$$D_2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}(H_{21}^{*}H_{21}+H_{22}^{*}H_{22})\cdot(R_{y11}(\omega)-\frac{\mid R_{y12}(\omega)\mid^2}{R_{y22}(\omega)})d\omega \tag{9}$$

A derivation of this formula must be omitted for lack of space.

Likewise we can find the distortion D_1 associated with the loss of $y_2[n]$. Since the two channels have the same erasure probability, the expected distortion due to erasure is:

$$D = \frac{1}{2}(D_1 + D_2). \tag{10}$$

In conclusion our optimization problem is to minimize D for a given redundancy ρ. Recall that the redundancy indirectly influences the distortion. It influences the matrix $T(\omega)$ which in turn influences $H(\omega)$ through (3) and $R_y(\omega)$ though (2).

III. Optimal solution

Before focusing our attention on the construction of the matrix $T(\omega)$, we want to remark some points about the decorrelating matrix $M(\omega)$.

A. Construction of $M(\omega)$

In the block transform context the decorrelating matrix (M) is clearly the Karhunen-Loève transform. If we further impose M to be unitary ($M^{+}M = I$) then we know that the solution is unique (orthonormal case), otherwise the solution is unique up to a scaling factor (orthogonal case). In the infinite dimensional case we are forced to work in the spectral domain and we have to diagonalize a p.s.d. matrix. It can be shown [11] that in case of a stationary input process the decorrelating matrix can be found analytically and it has the following shape:

$$M(\omega) = \begin{bmatrix} \frac{e^{j\omega/2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{e^{-j\omega/2}}{\sqrt{2}} \end{bmatrix}. \tag{11}$$

This matrix is clearly unitary and represents the solution we will use in the rest of the paper. It is important to point out that there are many singular cases where the solution to this problem is not unique, but even in such cases the proposed analytical solution exists.

B. Construction of $T(\omega)$

From now on we make the further hypothesis that $R_1 = R_2$ which is equivalent to requiring to transmit equal power over the two channels [1]. To develop our formulation we refer to the results published in [5]. Here Goyal et al. show that the optimal transform, in case of transmission of two Gaussian decorrelated variables over two independent channels with the same erasure probability and at the same rate ($R_1 = R_2$), is:

$$T = \begin{bmatrix} a & \frac{1}{2a} \\ -a & \frac{1}{2a} \end{bmatrix}, \tag{12}$$

where the value of a depends on the redundancy ρ:

$$a = \sqrt{\frac{\sigma_2}{2\sigma_1(2^\rho - \sqrt{2^{2\rho}-1})}}; \tag{13}$$

σ_1^2 and σ_2^2 are the variances of the two Gaussian components and the usual assumption is that $\sigma_1^2 > \sigma_2^2$. Finally the erasure distortion is given by:

$$D = \sigma_1^2 - \frac{1}{2\cdot 2^\rho(2^\rho - \sqrt{2^{2\rho}-1})}(\sigma_1^2 - \sigma_2^2). \tag{14}$$

It is interesting to notice that if the source has a circularly symmetric probability density, i.e., $\sigma_1 = \sigma_2$, then the distortion is independent of ρ. In fact in this case, because of the complete symmetry of the problem, it is completely useless to try to add correlation between the two input components.

We would like to generalize these results to our case, where the two variances vary with the frequency. Without loss of generality let us suppose that $\sigma_1^2(\omega) \geq \sigma_2^2(\omega)$ $\forall\omega$, a possible behavior of these variances could be the one depicted in Figure 5. Now we can divide the

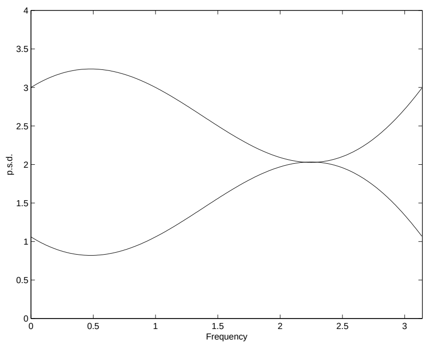

Fig. 5: The two p.s. distributions

frequency axis in N equal sub-intervals and, as a first approximation, suppose that the two power spectral densities are constant in each of these sub-intervals. (See Figure 6). Now we have only N possible values for $\sigma_1^2(\omega)$ and $\sigma_2^2(\omega)$

$$(\sigma_{1i}^2, \sigma_{2i}^2) \quad for \quad i = 1..N. \tag{15}$$

Thanks to this approximation we can apply Goyal's results on each interval and claim that the optimal transform for the generic ith intervals, given a redundancy ρ_i for that interval, is:

[1]This hypothesis, although reasonable, is not strictly necessary; but it simplifies the solution.

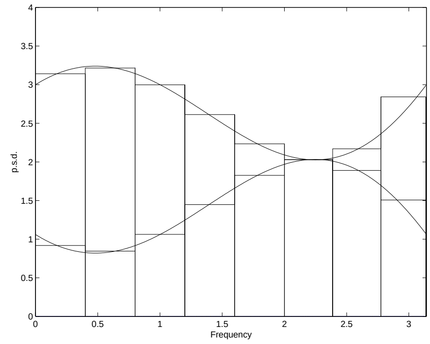

Fig. 6: The two approximated p.s. distributions

$$T_i = \begin{bmatrix} a_i & \frac{1}{2a_i} \\ -a_i & \frac{1}{2a_i} \end{bmatrix}, \tag{16}$$

where a_i is given by:

$$a_i = \sqrt{\frac{\sigma_{2i}}{2\sigma_{1i}(2^{\rho_i} - \sqrt{2^{2\rho_i} - 1})}}, \tag{17}$$

and that the erasure distortion is:

$$D_i = \sigma_{1i}^2 - \frac{1}{2 \cdot 2^{\rho_i}(2^{\rho_i} - \sqrt{2^{2\rho_i} - 1})}(\sigma_{1i}^2 - \sigma_{2i}^2). \tag{18}$$

However we want to minimize the global distortion:

$$D = \frac{1}{N}\sum_i D_i \tag{19}$$

given a global redundancy budget

$$\rho = \frac{1}{N}\sum_i \rho_i. \tag{20}$$

This is a typical problem of constrained minimization, so we define a new cost function L which combines the distortion and the redundancy through a positive Lagrange multiplier λ:

$$\begin{aligned} L &= D + \lambda\rho \\ L_i &= D_i + \lambda\rho_i \qquad\qquad \forall i = 1..N \end{aligned} \tag{21}$$

Finding a minimum of L (which now depends on λ too) amounts to finding minima for each L_i (because the costs are additive). Writing distortion as a function of the redundancy, $D_i(\rho_i)$, and taking the derivative we get:

$$\frac{\partial L_i}{\partial \rho_i} = \frac{\partial D_i}{\partial \rho_i} + \lambda = 0, \tag{22}$$

Thus, for a solution to be optimal, the set of chosen redundancy ρ_i has to correspond to constant-slope points on their respective redundancy-distortion curves. Uniqueness follows from the convexity of these curves and from the use of the Kuhn-Tucker conditions when necessary [10]. This problem formulation strongly recalls the problem of optimal bit allocation among a set of rate-distortion curves in order to minimize the total distortion represented by the sum of these single distortions [9, 7]. A constant-slope solution is obtained for any fixed value of ρ. To enforce the constraint (20) exactly, one has to search over all slopes λ until the budget is met. If we suppose that the redundancy budget is sufficiently large and that σ_{1i}^2 is never equal to σ_{2i}^2 then it is possible to give a closed form for the allocation problem. In fact:

$$\frac{\partial D_i}{\partial \rho_i} = -\frac{log(e)(\sigma_{1i}^2 - \sigma_{2i}^2)}{22^{\rho_i}(\sqrt{2^{2\rho_i} - 1})} \approx -log(e)(\sigma_{1i}^2 - \sigma_{2i}^2)2^{-2\rho_i} = -\lambda, \tag{23}$$

The constant-slope solution forces the redundancies to be of the following form:

$$\rho_i = \alpha + \frac{1}{2}log(\sigma_{1i}^2 - \sigma_{2i}^2). \tag{24}$$

Using the redundancy constraint (20) we can find α:

$$\alpha = \rho - \frac{1}{2N}\sum_i log(\sigma_{1i}^2 - \sigma_{2i}^2), \tag{25}$$

and finally

$$\rho_i = \rho + \frac{1}{2}log(\sigma_{1i}^2 - \sigma_{2i}^2) - \frac{1}{2N}\sum_i log(\sigma_{1i}^2 - \sigma_{2i}^2). \tag{26}$$

The approximation in (23) holds if ρ_i is sufficiently large. Its value depends of course on the total redundancy budget ρ, but also on the difference $\sigma_{1i}^2 - \sigma_{2i}^2$. If this difference is zero the corresponding side distortion (18) will not change with the redundancy and in this case it is better not to allocate any redundancy in this interval ($p_i = 0$). So when $\sigma_{1i}^2 = \sigma_{2i}^2$ the hypothesis of high redundancy budget is not enough to guarantee that the closed form (26) holds. In general we can say that the strategy of redundancy allocation on the different sub-intervals is strongly influenced by the difference $\sigma_{1i}^2 - \sigma_{2i}^2$, since the higher this difference is, the higher the corresponding redundancy will be.

Now that we know the optimal strategy of redundancy allocation, we can let the number N of intervals go to infinity (which means reduce the size of the intervals to zero) and find, in this way, the spectral optimal distribution of the redundancy for the two real p.s.d. $(\sigma_1^2(\omega), \sigma_2^2(\omega))$:

$$\rho(\omega) = \rho + \frac{1}{2}log(\sigma_1^2(\omega) - \sigma_2^2(\omega)) - \frac{1}{2\pi}\int_{-\pi}^{\pi} log(\sigma_1^2(\omega) - \sigma_2^2(\omega))d\omega. \tag{27}$$

The erasure distortion is:

$$D = \frac{1}{2\pi}\int_{-\pi}^{\pi} D(\omega)d\omega, \tag{28}$$

where

$$D(\omega) = \sigma_1^2(\omega) - \frac{1}{2 \cdot 2^{\rho(\omega)}(2^{\rho(\omega)} - \sqrt{2^{2\rho(\omega)} - 1})}(\sigma_1^2(\omega) - \sigma_2^2(\omega)). \tag{29}$$

The "recorrelating" transform $T(\omega)$ finally is:

$$\begin{bmatrix} a(\omega) & \frac{1}{2a(\omega)} \\ -a(\omega) & \frac{1}{2a(\omega)} \end{bmatrix}, \tag{30}$$

where:

$$a(\omega) = \sqrt{\frac{\sigma_2(\omega)}{2\sigma_1(\omega)(2^{\rho(\omega)} - \sqrt{2^{2\rho(\omega)} - 1})}}. \tag{31}$$

When the approximation (23) is not verified the only way to find the spectral distribution of the redundancy and of the other variables is via numerical computation.

Finally notice that the redundancy and the distortion definitions used in this section are consistent with the definitions given in the previous section. Strictly speaking the distortion used here refers to the mean square error between the decorrelated input sequences $\hat{x_1}, \hat{x_2}$ and the reconstructed sequences $\tilde{y}_1, \tilde{y}_2$ before entering the inverse of $M(\omega)$. Now this distortion represents also the end-system distortion since the matrix $M(\omega)$ is unitary.

IV. Application to first order Gauss-Markov processes

In this section we show our optimization results for a first order Gauss-Markov or Gauss autoregressive source $x[n]$ defined by:

$$x[n] = ax[n-1] + w[n], \tag{32}$$

where the regression coefficient a has magnitude less than 1 and where $w[n]$ is a zero mean, unit variance, i.i.d. Gaussian source. The input autocorrelation of this process is:

$$r_x[n] = \frac{a^{|n|}}{1 - a^2}. \tag{33}$$

Now, after downsampling, the two polyphase subsequences $x_1[n], x_2[n]$ are still Gauss-Markov processes, but with the regression coefficient a replaced by a^2 and the i.i.d. original Gaussian source ($w[n]$) replaced by a new i.i.d. Gaussian source with zero mean and variance $1 + a^2$. Hence the autocorrelation for this two processes is:

$$r_{x1}[n] = r_{x2}[n] = \frac{a^{|2n|}(1 + a^2)}{1 - a^4}, \tag{34}$$

and the corresponding p.s.d are:

$$R_{x11}(\omega) = R_{x22}(\omega) = \frac{1 + a^2}{|1 - a^2 e^{-j\omega}|^2} \tag{35}$$

To find $R_{x12}(\omega)$, we need to compute the cross-correlation $r_{x1x2}[n]$ first:

$$r_{x1x2}[n] = E(x_1[n]x_2[0]) = E(x[2n]x[1]) = \frac{a^{|2n-1|}}{1 - a^2}. \tag{36}$$

and, applying the Fourier transform:

$$R_{x12}(\omega) = \frac{a(1 + e^{(-j\omega)})}{1 + a^2} R_{x11}(\omega), \tag{37}$$

and $R_{x12}(\omega) = R^*_{x21}(\omega)$.

As already shown, the corresponding decorrelation matrix is:

$$M(\omega) = \begin{bmatrix} \frac{e^{j\omega/2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{e^{-j\omega/2}}{\sqrt{2}} \end{bmatrix}. \tag{38}$$

The p.s.d input matrix after decorrelation is the following:

$$R_{\hat{x}}(\omega) = \begin{bmatrix} R_{x11}(\omega)(1 + \frac{2acos(\omega/2)}{(1+a^2)}) & 0 \\ 0 & R_{x11}(\omega)(1 - \frac{2acos(\omega/2)}{(1+a^2)}) \end{bmatrix} \tag{39}$$

The behavior of these two p.s.d for $a = 0.5$ is depicted in Figure 7. As it can be noticed the two functions are equal only in π (and of course in $-\pi$).

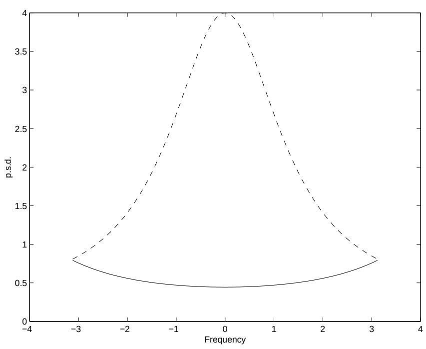

Fig. 7: The two p.s.d distributions after decorrelation for a Gauss Markov source with $a = 0.5$.

The next step consists in constructing the matrix $T(\omega)$. As we already said, in the points around the frequency values where $\sigma_1^2(\omega) = \sigma_2^2(\omega)$ it is not possible to use the closed-form (26) even in the high redundancy hypothesis. So, for the Gauss-Markov source, $a(\omega)$ (and so $T(\omega)$) can only be found numerically. In Figure 8 the behavior of $a(\omega)$ for $\rho = 2$ is shown. The final polyphase matrix $H(\omega) = T(\omega)M(\omega)$ is the following:

$$\begin{bmatrix} a(\omega)\frac{e^{(j\omega/2)}}{\sqrt{2}} - \frac{1}{2\sqrt{2}a(\omega)} & \frac{a(\omega)}{\sqrt{2}} + \frac{e^{(-j\omega/2)}}{2\sqrt{2}a(\omega)} \\ -a(\omega)\frac{e^{(j\omega/2)}}{\sqrt{2}} - \frac{1}{2\sqrt{2}a(\omega)} & -\frac{a(\omega)}{\sqrt{2}} + \frac{e^{(-j\omega/2)}}{2\sqrt{2}a(\omega)} \end{bmatrix} \tag{40}$$

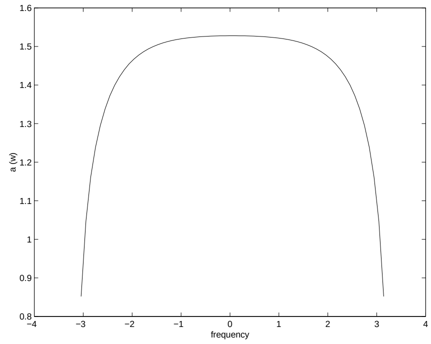

Fig. 8: $a(\omega)$ for $\rho = 2$ and $a = 0.5$.

The filters that exactly represent this polyphase matrix are of infinite length, but in practical settings, where infinite delay or complexity is not allowed, we have to look for FIR solutions. For this reason we approximated the ideal filters with finite length ones. Because of this approximation the erasure distortion will be clearly worse than the ideal, so we construct the FIR filters trying to minimize the square difference between the optimal distortion and the approximated one, keeping the constraint that the obtained filter banks verify the perfect reconstruction conditions ([7]). In Fig. 9 we show the ideal distortion and two approximated distortions for the case where all the filters have length $N = 4$ or length $N = 8$. We can notice that for both cases the behavior is quite close to the ideal one.

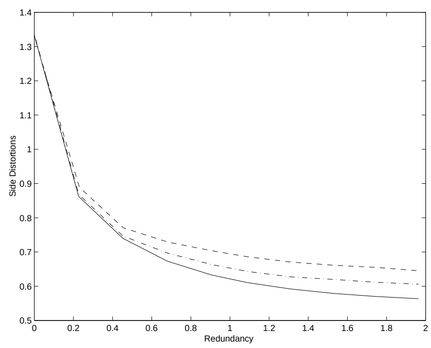

Fig. 9: The ideal and the two approximated distortions. Solid: ideal curve, dash-dotted: curve with FIR filters of length N=8, dashed: curve with FIR filters of length N=4.

V. Conclusion

In this work we have addressed the MDC problem in the transform subband coding context. We have generalized the results given in some previous papers that apply only on finite length input vectors to the more general and realistic case of input sequences and subband decomposition. We have shown a way to construct the ideal redundancy-distortion curve and the equivalent filter banks that can meet the points of this curve. We have also proposed a method to approximate the ideal filters with FIR filters. The results of this approximation on a first order Gauss-Markov source are quite encouraging. We are currently interested in seeing if these results are applicable to more realistic signals like speech or images.

Acknowledgments

Pier Luigi Dragotti was supported by Fonds National Suisse project number 21-52439.97

VI. References

[1] A.A. El Gamal and T.M. Cover, "Achievable rates for multiple descriptions," IEEE Trans. Information Theory, vol IT-28, no. 6, pp.851-857, Nov. 1982.

[2] L.Ozarow, "On a source coding problem with two channel and three receivers". Bell Syst. Tech. J., 59(10): pp. 1909-1921, December 1980.

[3] V.A. Vaishampayan, "Design of multiple description scalar quantizers," IEEE Trans. Information Theory,vol.39, no.3, pp.821-834, May 1993.

[4] M.T. Orchard, Y. Wang, V.Vaishampayan, and A.R.Reibman, "Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms," In Proc. IEEE Int. Conf. Image Proc., Santa Barbara, October 1997.

[5] V. Goyal and J.Kovacevic, "Optimal multiple description transform coding of Gaussian Vectors", In Proc. Data Compression Conference, Snowbird, March 1998.

[6] Xuguang Yang, Kannan Ramchandran, "Optimal multiple description subband coding," In Proc. IEEE Int. Conf. Image Proc., Chicago, October 1998.

[7] M.Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice Hall, 1995.

[8] T.Berger, *Rate Distortion Theory*, Prentice-Hall,Englewood Cliffs, NJ, 1971.

[9] Y.Shoham and A.Gersho. "Efficent bit allocation for an arbitrary set of quantizers," IEEE Trans. Acoust. Speech, and Signal Proc., 36 (9) pp.1445-1453,September 1988.

[10] E.K.P. Chong and S.H. Zak *An introduction to Optimization* John Wiley & Sons 1996

[11] M.K.Tsatsanis and G.B. Giannakis, "Principal Component Filter Banks for Optimal Multiresolution Analysis," IEEE Trans. Sig. Proc., Vol.43, No.8, pp.1766-1777, Aug.1995